4 Statistics and probability

Teaching support and guidance

Concepts

- Approximation
- Modelling
- Systems

Outcomes

Students will understand how statistical representations and measures allow us to represent data in many different forms to aid interpretation. Both statistics and probability provide important representations that enable us to make predictions, valid comparisons and informed decisions.

Conceptual Understandings

- Organising, representing, analysing and interpreting data, and using different statistical tools, facilitate prediction and drawing of conclusions.
- Approximation in data can approach the truth but may not always achieve it.
- Variance, or change, in probability density functions can be identified as a measure of central tendency, such as mean, mode and median.

Inquiry Questions

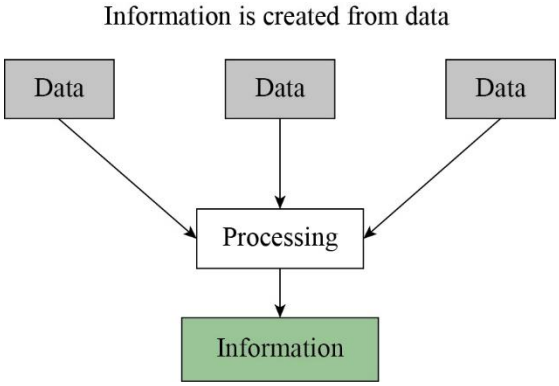
- Factual: What is the difference between information and data?
- Conceptual: How do we misuse statistics?
- Debatable: Can all data be modelled by a function?

Factual: What is the difference between information and data?

Concept: Approximation

Standard Level

PowerPoint: Sampling techniques (S4.1)



Further Questions:

What process (or processes) does data go through to become information? How can we be sure our data is valid?

Discuss with students the process of sampling. The PowerPoint outlines some of the principal sampling techniques you may wish to use, with their advantages and disadvantages, as well as some possible applications of each one.

They are:

- random sampling
- systematic sampling
- convenience sampling
- stratified sampling and quota sampling.

Have students think about the following question:

- How can we be sure that a random sample is truly random?

Conceptual: How do we misuse statistics?

Concepts: Approximation, Systems

Standard Level

Discussion: International mindedness

The benefits of sharing and analysing data from different countries; discussion of the different formulae for variance.

PowerPoint: Variance (S4.3)

Emphasize that, generally, appropriate technology should be used to calculate variance. However, there is an argument that manual calculation can improve understanding of the relation of variance to the actual data. The PowerPoint offers a scaffolded approach to the calculation. This is useful if any students need to make the calculation for an exploration.

Population variance versus sample variance

$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$ OR $s^2 = \frac{\sum(x-\mu)^2}{N-1}$

When we undertake a project, we almost never have access to data for an entire population. For example, we might be able to measure the height of everyone in the classroom, but we cannot measure the height of everyone on Earth. Our data is only a sample of the entire population. This means we must use a slightly different formula to calculate variance, with an $N-1$ term in the denominator, instead of N.

This could lead into a discussion based on the conceptual question.

Debatable: Can all data be modelled by a function?

Concepts: Modelling, Approximation

Standard Level

PowerPoint: Calculation of the r-value (S4.4)

Emphasize that students should use appropriate technology to calculate r. However, there is an argument that the manual calculation can enhance understanding of the relation to the actual data. The PowerPoint offers a scaffolded approach to the calculation. It is useful if any students need to make the calculation for an exploration.

Spreadsheet: Normal distribution function (S4.8)

The spreadsheet is designed as an exploration piece, allowing students to experiment with the normal distribution function. It may also be useful for students who choose to focus their exploration piece on the topic of probability.

Spreadsheet: Binomial distribution (S4.9)

The spreadsheet is designed as an exploration piece, allowing students to experiment by using different numbers of outcomes and probabilities. It may also be useful students who choose to focus their exploration piece on the binomial distribution.

Links: Binomial distribution discussion

Discussion points regarding using the binomial distribution in modelling are offered on the website:

- www.mathsisfun.com/data/binomial-distribution.html

Again, this will be useful as an introductory tool or for students who choose an exploration piece based on this topic.

Higher Level

Activity: The Monty Hall problem investigation (H4.13)

This is an introductory group activity that encourages discussion about Bayes' theorem. The teacher version includes the calculations based on Bayes' theorem; however, it may be useful to ask students to perform the calculations as they complete the task. You will need to print out enough cards for each pair of students to have a set of three.

A discussion of the use of Bayes' theorem to simulate the Monty Hall problem is available on:

- https://medium.com/@NickDoesData/applying-bayes-theorem-simulating-the-monty-hall-problem-with-python-5054976d1fb5

Links: Bayes' theorem (H4.13)

Bayes' theorem at brilliant.org offers discussion points regarding using Bayes' theorem in modelling via this website:

- https://brilliant.org/wiki/bayes-theorem

Again, this will be useful as an introductory tool or for students who choose an exploration piece based on this topic.